

University of Alberta Library



0 1620 3369973 5

For Reference

NOT TO BE TAKEN FROM THIS ROOM

Ex LIBRIS
UNIVERSITATIS
ALBERTAEensis





Digitized by the Internet Archive
in 2017 with funding from
University of Alberta Libraries

<https://archive.org/details/relationshipbetw00kost>

THE UNIVERSITY OF ALBERTA

RELATIONSHIP BETWEEN SCORES OBTAINED
BY BENNETT AND BERNREUTER SYSTEMS
OF SCORING BERNREUTER INVENTORY

A DISSERTATION SUBMITTED
TO THE COMMITTEE ON GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE DEGREE OF
MASTER OF EDUCATION

FACULTY OF EDUCATION

BY

WILLIAM E. KOSTASH

EDMONTON, ALBERTA

JULY, 1949

TABLE OF CONTENTS

	Page
LIST OF TABLES	iii
CHAPTER	
I. INTRODUCTION	
1. History of the Problem	1
2. Bennett Revision	3
II. THE PROBLEM	
1. Purpose of Investigation	6
2. Material of Investigation	7
III. EXPERIMENTAL DESIGN	9
IV. DATA OBTAINED	10
V. SIGNIFICANCE OF DATA OBTAINED	
1. Reliability of Obtained Coefficients ..	13
2. Significance of Obtained Coefficients .	15
3. Significance of Differences between Obtained Coefficients	16
VI. SUMMARY AND CONCLUSION	19

LIST OF TABLES

		PAGE
TABLE I	DISTRIBUTION OF INVENTORIES	8
TABLE II	COEFFICIENTS OF CORRELATION BETWEEN THE PERCENTILE SCORES OBTAINED BY THE TWO METHODS	10
TABLE III	COEFFICIENTS OF CORRELATION OBTAINED BY PRODUCT-MOMENT AND BY AVERAGING FISHER Z-FUNCTIONS	11
TABLE IV	COEFFICIENTS OF CORRELATION CALCULATED BY AVERAGING CORRESPONDING Z-FUNCTIONS	12
TABLE V	COMPARISON OF COEFFICIENTS OF CORRELATION OBTAINED IN THIS INVESTIGATION AND BY BENNETT	19

CHAPTER ONE

INTRODUCTION

1. HISTORY OF THE PROBLEM

Bernreuter's Personality Inventory, as a means of measuring personality, is only one of many instruments that have been developed in recent years in the field of personality studies. It is, however, the best known and most widely used today, and has features that make it particularly well-adapted to situations where it is desired that certain personality abnormalities be detected for guidance or psychiatric treatment. As a group test it may be administered to many persons at one time, and it has the further advantage over many similar tests in that it canvasses several different aspects of personality at once, in this way effecting considerable saving both in cost and in the time required for administration.

The Personality blank itself contains 128 questions which may be answered: Yes, No, ? It may be scored on each of six scales. Four of them (the B scales), prepared by Bernreuter and designated by symbols B1-N, B2-S, B3-I, and B4-D, may be described as follows:

B1-N. A measure of neurotic tendencies.

B2-S. A measure of self-sufficiency.

B3-I. A measure of extroversion-introversion.

B4-D. A measure of dominance-submission.

Separate scoring keys are used in scoring the inventory blank, one for each trait tested. In the preparation of these keys the diagnostic value of each response to each question was determined for each of the traits. Weights from +7 to -7 were assigned in accordance with these diagnostic values. The total score for a trait is the algebraic sum of the weights which correspond to the responses made by the subject, as given on the key, for that particular trait.

The addition of the two F scales, F1-C and F2-S, came about in the development of the Inventory. Bernreuter found that neurotic tendency and introversion-extroversion correlated about .95, so that they were virtually the same. Thus only three of the original four B scales were needed. Flanagan, by means of factor analysis, drew the conclusion that two traits, which he named self-confidence and sociality, were the chief components of the first four. These have been built into the scoring keys, F1-C and F2-S, and can be used in place of the other four. Thus the inventory blank may be scored on the two F scales only, or on all six, depending on the preference of the administrator.

Several features of the Bernreuter and Flanagan scales give rise to serious disadvantages in the use of the inventories. In the first place, the scoring of the six scales is a tedious affair. Except when the Veedor Counter, or the I.B.Machine, is used, hand scoring the six scales requires about thirty

minutes for each blank. Second, since the F scales appear to give equally reliable information as the individual scores on the four B scales, it seems of dubious worth to use the B scales at all. Finally, Bernreuter himself admits that "the inter-correlations between Bl-N, B3-I, and Fl-C scales are very high. Little is gained through using more than one of these three scales".

These features of the B and F scales prompted numerous attempts at revision and simplification of the original Bernreuter method of scoring. The one with which this dissertation is concerned was put forward by Geo. K. Bennett.

2. BENNETT REVISION

Bennett describes his method of revision and simplification as follows:

"It is apparent that if the blank could be scored for only the two Flanagan scales, the weights of these scores reduced to small positive numbers, and the location of percentile equivalents simplified, much time would be saved in scoring. Accordingly, the following steps were taken:

1. A tabulation was made of the range of weights assigned to each item in the two F scales.
2. Equivalent small positive weights were assigned according to the original total weights."

Bennett then prepared scoring keys for each of the two Flanagan scales by means of which raw scores could be obtained on the basis of the new weights assigned to each item in the blank. Since the new weights ranged from 0 to +2, the scoring by the Bennett method effected a tremendous saving of time.

He further found that the coefficients of correlation between the raw scores obtained by the original method of scoring and by this revised and simplified method were .971 for F1-C and .938 for F2-S. These high correlations seemed to justify the use of the simplified method of scoring the two Flanagan scales.

Furthermore, since the F scales could be used to estimate the scores on the four B scales, Bennett compared the scores obtained on the four B scales by means of the Bernreuter keys with scores estimated by means of the simplified keys, and found the following correlations:

B1-N -- .95

B3-I -- .951

B2-S -- .934

B4-D -- 889

To make the raw scores obtained on the four B scales more meaningful, Bernreuter converted them to percentiles. Similarly, Bennett obtained percentile equivalents for scores obtained by his simplified method. He then obtained two sets of percentile equivalents for a group of adult women whose blanks were scored by the Bernreuter keys and by the simplified

fied keys. Comparing the two sets of percentiles thus obtained, he found the following correlations:

B1-N -- .951

B3-I -- .934

B2-S -- .883

B4-D -- .924

Bennett summarizes his findings as follows:

"Since the multiple correlation is less than unity, there is some loss of accuracy by this method

In view of the fact that the traits which we are attempting to measure are subject to large quotidian fluctuations, at least as measured by the original scales, it is believed that the economy of scoring more than justifies the small loss in precision."

REFERENCES

1. Bennett, Geo. K. - A Simplified Scoring Method for the Bernreuter Personality Inventory.
2. Mursell, J. L. - Psychological Testing.

CHAPTER TWO

THE PROBLEM

1. PURPOSE OF INVESTIGATION

It is demonstrably true that the Bennett simplified system of scoring the Bernreuter Personality Inventory blanks effects considerable economies of time and effort. If it can be confirmed that this system also yields substantially the same results as the original Bernreuter method, then there would appear to be ample justification for the use of the simplified in place of the original system of scoring the Bernreuter Inventory blanks.

It has already been pointed out that Bennett submitted data to prove that correlation between the results obtained by the two methods is sufficiently high to warrant the use of his simplified system in the interests of economy. However, he based his contention on statistics he obtained from research which he carried out on relatively small samples drawn from populations found in American cities. For example, to obtain the coefficients of correlation between the original and simplified Flanagan scales, he used 115 Personality Inventory blanks. To obtain the statistics for his chart of Bernreuter Percentile Equivalents for the Bennett simplified scoring system, he used the blanks scored by 100 adult women.

Those who use the Bernreuter Personality Inventories in Canada feel a need for some evidence that the use of the Bennett simplified scoring system is justified on the same grounds as

those claimed by Bennett. This dissertation and investigation, therefore, has a two-fold aim:

1. to determine statistically to what extent the Bennett system of scoring the Bernreuter Inventory blanks yields substantially the same results as those obtained by the original Bernreuter method, when the blanks are administered to groups selected in Canada.
2. to compare the results obtained in this investigation with those claimed by Bennett.

2. MATERIAL OF INVESTIGATION

Both Bennett and Bernreuter computed separate percentile norms for the three main groups for which the inventory blanks were primarily devised, and for both sexes of each group; viz., college men and women, adult men and women, and high school boys and girls.

This investigation concerns itself with the first two groups only; and the material which furnished the basic data therefore was made up of one hundred and ninety blanks completed by:

1. forty adult men
2. fifty adult women
3. fifty college men
4. fifty college women

The blanks were administered to male and female members of the Faculty of Education in Edmonton, and male and female members of the staffs of radio stations in several cities in Canada. Table I shows the distribution of the inventories among the four groups:

TABLE I
DISTRIBUTION OF INVENTORIES

	Adult Men	Adult Women	College Men	College Women
<u>No Occupation given</u>	13	14		
<u>Employed by Radio Station:</u>				
CKWX	18	11		
CKRC	13	13		
CKLN	1			
CJAT	3	3		
CJVI	2	3		
CKOC		3		
CJCA		3		
<u>Faculty of Education, University of Alberta</u>			40	50
Totals	50	50	40	50

CHAPTER THREE

EXPERIMENTAL DESIGN

The first step in the investigation consisted of scoring the inventories by the Bernreuter method. Because each blank yields raw scores for each of the four B scales -- B1-N, B2-S, B3-I, and B4-D, it had to be scored four times. This consumed a great deal of time, and required over one hundred hours for the entire one hundred and ninety blanks.

The second step consisted of transmuting, by means of the Tentative Percentile Norms, the raw scores to percentile equivalents. This was done for each of the four B scales for the one hundred and ninety inventories. This part of the process consumed another fifty hours.

In the third step, the same hundred and ninety blanks were scored, on the F scales, by the Bennett Simplified Scoring system. For these F scales, Bernreuter percentile equivalents were read off from the Bennett Percentile Equivalent chart, and entered on each blank below the percentiles obtained from the Bernreuter Tentative Norms chart. By contrast, this step required less than 25 hours.

The final step was the running off of coefficients of correlation between the two sets of percentiles obtained for each personality trait measured and for each sub-group of the sample. These coefficients of correlation are shown in Table III.

CHAPTER FOUR

DATA OBTAINED

In calculating the coefficient of correlation between the percentiles obtained by the two methods, viz., the original Bernreuter and the Bennett Simplified systems, the product-moment method was used. This method of calculation was decided upon for two reasons. In the first place it lends itself to ease of computation in that the data can be grouped and presented in tabular form, and the coefficient of correlation, r , calculated by the short method. In the second place, the sample is sufficiently large to warrant the use of this method. In calculating r between the scores on the four B scales made by the four sub-groups, 760 raw scores were compared.

TABLE II

COEFFICIENTS OF CORRELATION BETWEEN THE PERCENTILE SCORES OBTAINED BY THE TWO METHODS

Sub-group	N	B1-N	B2-S	B3-I	B4-D	All traits
College Men	40	.96	.93	.95	.88	.93
College Women	50	.96	.89	.88	.89	.90
Adult Women	50	.96	.93	.92	.92	.92
Adult Men	50	.98	.92	.97	.89	.90
All Groups	190	.96	.92	.93	.90	
Total Group	760					.91

As a check on the calculation of the coefficient of correlation by the Product-Moment method, the r 's, obtained for the individual sub-groups and tendencies measured, were averaged by calculating the Arithmetic Mean of the corresponding Fisher Z-Functions₁, and the two sets of coefficients compared. Table III shows the comparison.

TABLE III

COEFFICIENTS OF CORRELATION OBTAINED BY PRODUCT MOMENT-METHOD AND BY AVERAGING FISHER Z-FUNCTIONS

	Product-Moment(r)	Averaging X-function
College Men	.93	.93
College Women	.90	.91
Adult Women	.92	.93
Adult Men	.90	.95
B1-N	.96	.96
B2-S	.92	.92
B3-I	.93	.94
B4-D	.90	.89
Total Group	.91	.92

1. Garrett, H.E.: Statistics in Psychology and Education, page 302

The fact that almost the same r was obtained for all the groups and traits by the much shorter method of averaging the Z-functions seemed to warrant its use in arriving at other coefficients of correlation. Table IV shows these other correlations.

TABLE IV

COEFFICIENTS OF CORRELATION CALCULATED BY AVERAGING CORRESPONDING Z-FUNCTIONS

Sub-group	r
College	.91
Non-college	.91
Male	.91
Female	.91

To sum up, twenty-five r 's were calculated by the product-moment method, four by averaging Fisher Z-functions corresponding to relevant r 's arrived at by the first method, and sixteen by both methods.

References

1. Garrett, H.E. - Statistics in Psychology and Education.
2. Lindquist, E. F. - Statistical Analysis in Educational Research.
3. Peatman, J. S. - Descriptive and Sampling Statistics.

CHAPTER FIVE

SIGNIFICANCE OF DATA OBTAINED

1. RELIABILITY OF OBTAINED COEFFICIENTS

Table II shows coefficients of correlation ranging from .88 to .98. These are high correlations, but are they reliable? Do they signify that both methods of scoring the Bernreuter Personality blanks yield substantially the same percentiles? Or could such high correlations be obtained by errors in sampling when, in fact, the true or parameter correlation between the scores obtained by the two methods is zero?

The reliability of the obtained r 's was tested, first, against the null hypothesis. On the hypothesis that the correlation between the scores obtained by the two methods of scoring the Bernreuter blanks is zero, it is known that an r of .361, when $N = 50$, can be obtained by accidents of sampling. The r 's ranging from .88 to .98, make this hypothesis untenable, for correlations as high as .88 or .98, at the .01 level of significance, cannot be attributed to chance alone. That is to say, the probability that an obtained r could be as high as .88, if the true r is zero, is much less than one in a hundred.

It may be inferred, therefore, that such high correlations are significant. In this thesis it is assumed that the construction of the Bennett Simplified scoring key is such

that percentiles obtained by its use correlate positively and highly with scores obtained when the same inventories are scored by the original Bernreuter method.

The reliability of the obtained r 's was further tested by computing the limiting values of the corresponding parameter coefficients of correlation (r_h). For an r of .90, when $N = 50$, using the Fisher Z-function formula₁, the limiting values of r_h are .80 and .95. The obtained r of .90 lies within this range.

When N is greater the limits of r_h are narrower still. For instance, in computing the r between the percentiles obtained by the two methods of scoring the inventories on all four B scales, 760 percentiles are compared. The limits of r_h are .89 and .92. The obtained r is .91, virtually half-way between the narrow limits ascertained for the true r . It is wholly conceivable that if several thousand percentile scores were to be compared, the true r and the obtained r would be identical. The close correspondence between the parameter r and the obtained r , especially for higher values of N , seems to justify the inference that the latter is highly reliable.

Thus it is not only disproved that an r as high as .91, when $N = 760$, can be obtained when the true r is zero, but it is also shown that the obtained r falls within the narrow limits ascertained for the true r .

1. Lindquist, E.F.: Statistical Analysis in Educational Research, Table XIII, page 212.

2. SIGNIFICANCE OF OBTAINED COEFFICIENTS

Having demonstrated the high reliability of the obtained r 's, the question arises: Are the coefficients significant? If an r as high as .361, when $N = 50_1$, can be obtained by a sheer accident of sampling, when in reality no correlation exists, can it be taken that r 's ranging from .88 to .98 signify some intrinsic feature of the Bennett Simplified system to yield, in a high proportion of cases, the same scores as the Bennett system? It has already been asserted that that is the inference₂. It can further be substantiated that an r of .90 has an 80% dependability.

Garrett developed a formula by which he computes the coefficient of forecasting dependability. He designates the coefficient by the symbol E . The formula is

$$E = 1 - \sqrt{1 - r^2}$$

Table II shows that the coefficient of correlation between the scores on the Bl-N test (adult men) is .98. Applying the formula for dependability we find that $E = .80$, and the efficiency in predicting scores may be said to be 80%. That is, we can be certain eight times out of ten that the same percentile score can be obtained when the Bl-N test is scored by either the original Bernreuter method or by the Simplified Bennett system. Even an r as low as .88 has a predictive value of 50%; that is, halfway between no predictive value and perfect prediction₃.

1. Lindquist, E.F.: Statistical Analysis in Educational Research, Table XIII, page 212.
2. See page 13.
3. Garrett, H.E.: Statistics in Psychology and Education, page 337.

3. SIGNIFICANCE OF DIFFERENCES BETWEEN OBTAINED COEFFICIENTS

Some differences between obtained r 's were obtained. For instance, the coefficients of correlation on the B3-I scale range from .88 for College Women to .97 for Adult Men. Since the same test, B3-I, was applied to independent samples, does it mean that one cannot rely with the same degree of certainty that the scores for the college women will correlate as highly as they do for adult men? Or does it mean that this difference may be attributed to errors in sampling, and that, in reality no difference exists? To answer these questions, this difference, as well as others, was tested for significance.

The significance of the difference between the r 's obtained when the B3-I scale was scored by the two methods for both College Women and Adult Men was tested first. Both samples contain 50 items. The coefficient of correlation (r_1) for the first sample, College Women, is .88. The r for the other sample (r_2), Adult Men, is .97. Converted to Fisher Z-functions¹ when $r_2 = .97$, $Z_2 = 2.09$; and when $r_1 = .88$, $Z_1 = 1.38$.

$$Z_2 - Z_1 = 2.09 - 1.38 = .71$$

The standard error of this difference is .21.

1. Peatman, J.S.: Descriptive and Sampling Statistics,
Table 13:1, page 386.

The critical, or test ratio, T , at the .01 level, is 2.56. That is, the hypothesis that the true difference between the two coefficients of correlation is zero is tenable when T is 2.56 or less. T is obtained by dividing the difference between the two r 's by the standard error of this difference, the r 's being converted to Z -functions, as above.

T , therefore, is $.71/.21$ or 3.4. Since it is greater than 2.56, one might be tempted to reject the null hypothesis, and to conclude that the difference between the two r 's is not a chance difference, nor the true or parameter difference actually zero and, therefore, of no significance.

In other words, one might feel justified in concluding that, through some defect, the Bennett Simplified system of scoring the B3-I scale cannot be reliably used when the Bernreuter inventory is administered to adult men or to college women. Such a conclusion, however, is not justified. It cannot be stated with confidence that the correlation between the percentile scores of the adult men is higher than that between those scored by college women. The obtained r for the first group is between .94 and .98 which are the limits of the true r for that group; while the obtained r for the second is between .76 and .94. The true r for the first may be as low as .94, and the true r for the second as high as .94. In short, theoretically, the two may be the same. Consequently, the apparent difference may not be great enough to be significant. In terms of the null hypothesis, one cannot reject the possibility that the true difference is zero.

All the other differences between the obtained r's were also tested and found, in terms of T, to be insignificant. It seems, therefore, that in fact there are no grounds for holding that any real differences between the obtained r's exist. To put it positively, the two methods of scoring the Bernreuter inventories completed by any of the four sub-groups yield the same results, and may, therefore, be used interchangeably.

REFERENCES

Peatman, J. S.: Descriptive and Sampling Statistics.

CHAPTER SIX

SUMMARY AND CONCLUSION

The findings of this investigation may now be summarized.

Reliably high coefficients of correlation, ranging from .88 to .98, between percentile scores obtained by the Bernreuter and the Bennett Simplified systems of scoring Bernreuter Personality blanks were calculated by the Product-Moment (short) method.

Table V shows the coefficients of correlation obtained in this investigation and those obtained by Bennett.

TABLE V
COMPARISON OF COEFFICIENTS OF CORRELATION OBTAINED IN THIS INVESTIGATION AND BY BENNETT

	IN THIS INVESTIGATION (N = 190)	BY BENNETT (N = 100)
B1-N	.96	.951
B2-S	.92	.893
B3-I	.93	.934
B4-D	.90	.924

When the differences between the r's in Table V were tested, none was found to be significant.

From the foregoing summary of findings it may be concluded that

- a. The Bernreuter and Bennett Systems of scoring the Bernreuter Inventories are interchangeable to a high degree.
- b. For the sake of economy of time and effort, the Bennett system may be used in preference to the Bernreuter.
- c. The use of the Bennett Simplified system of scoring Bernreuter Inventories completed by groups of Canadian adults and College men and women is warranted.

B29757